

**A Digital Challenge: Bringing Kappler's *Indian Affairs: Laws and Treaties* to the
World Wide Web**

**Amigos Fellowship Final Report
January 1998**

Submitted by

**Suzanne Holcomb, Principal Investigator
Oklahoma State University
The University Library**

Indian Affairs: Laws and Treaties.
Vol. II (Treaties) in part. Compiled and edited by Charles J. Kappler .
Washington: Government Printing Office, 1904.

[Home](#) | [Table of Contents](#) | [Index](#) | [Search](#)

INTRODUCTION

A Digital Challenge: Bringing Kappler's *Indian Affairs: Laws and Treaties* to the World Wide Web. Report Sununary.

Oklahoma State University Library
December 1997

Table of Contents

[Overview](#)

[Text](#)

[Objectives](#)

[Equipment](#)

[Process and Format](#)

[Evaluation](#)

[Conclusion](#)

[Project Participants](#)

[Questions and Comments](#)

Overview

In May 1996, the AMIGOS Bibliographic Council awarded the Oklahoma State University Library a \$1,500 grant to convert 150 pages of complex text from Volume II (Treaties) of *Indian Affairs: Laws and Treaties* compiled and edited by Charles J. Kappler to a digital format, maintaining as much as possible the appearance and intent of the original work while allowing for enhanced access including full-text indexing. These pages include all of the pre-removal treaties of the Five Tribes: Cherokee, Chickasaw, Choctaw, Creek and Seminole. These tribes were chosen because of their significance to the state of Oklahoma.

[Table of Contents](#)

Text

Indian Affairs: Laws and Treaties, compiled and edited by Charles J. Kappler, is an historically significant, seven volume compilation of U.S. treaties, laws and executive orders pertaining to Native American Indian tribes. The volumes cover U.S. Government treaties with Native Americans from 1778-1883 (Volume II) and U.S. laws and executive orders concerning Native Americans from 1871-1970 (Volumes I, III - VII). The work was first published in 1903-04 by the U.S. Government Printing Office. Enhanced by the editor's use of margin notations and a comprehensive index, the information contained in *Indian Affairs: Laws and Treaties* is in high demand by Native peoples, researchers, journalists, attorneys, legislators, teachers and others of both Native and non-Native origins.

[Table of Contents](#)

Objectives

The objectives of the project were:

- To develop an innovative methodology for reproducing on the WWW text containing complex formats including margin annotations, variable type fonts and tabulated and columnar information
- To use hypertext and full-text indexing to enhance historical or legal documents;
- To determine the cost effectiveness of the developed methodology to digitize the complete set of Kappler's *Indian Affairs: Laws and Treaties* as well as other historical or legal documents;
- To make available on the WWW a faithful representation of an historically significant and long out-of-print resource to the benefit of Native Americans, historians, students and the legal community;
- To enhance the skills and knowledge of the project participants relative to presenting information on the WWW; and,
- To attract additional private funding based on the strength of the successfully completed project.

Table of Contents

Equipment

The Library used its Hewlett-Packard Scanjet 4C scanner running Caere's Onnipage Pro software v6.0 for scanning the pages using Optical Character Recognition (OCR). Image Assistant, for scanning images, was also installed. Initial editing was done in Onnipage; detailed editing was completed in Microsoft Word. The project files are on the Library's server, running Microsoft's Internet Information Server (IIS) v4.0 for serving the Web pages and the included Microsoft Index Server v2.0 for indexing the pages.

Table of Contents

Process and Format

As the project goal was to publish text from *Indian Affairs: Laws and Treaties* on the Web, replicating the original text while allowing for enhanced access including full-text indexing, the project group used Onnipage Pro software to scan the text using OCR, resulting in searchable text. The text was then tagged in HyperText Markup Language. Two students were hired to do the scanning, editing and proofreading of 58 treaties, representing 150 pages of text. The staff completed the final proofreading and the HTML markup.

Copies of the original text were made and enlarged by 20% to enhance the performance of the OCR due to the font and font size of the original text. The most efficient way to zone the text for scanning was to zone the heading, page numbers, entire text (excluding the side notes) and the signatures as two separate zones if in two columns. If the signatures were lengthy, several zones within one column had to be used. If the side notes were close together, they could be scanned as one zone and the sentences then separated. If they were far apart, as was most usual, they were not zoned and re-keyed later. Onnipage allowed for initial editing. Detailed editing and proofreading took place following this. All tagging was done in Notepad (or Microsoft Word with larger files) due to the occasional problems with HTML editors.

Two versions of each treaty are available. A high-end version uses tables to replicate *Indian Affairs: Laws and Treaties* as closely as possible, permitting the use of current Web techniques and serving advanced users. A text-only version is available for users with limited hardware and software (Lynx users, users with a voice-reader, etc.). Web accessibility for all users is a key issue.

Links for Home, Introduction, Table of Contents, Search, Index and Pages are at the top of each treaty in both formats. The table of contents and search function links are value-added capabilities: the table of contents is organized by tribe and the search function features Microsoft's Index Server v2.0. In addition, Kappler's notes are hyperlinked and at the top left (table) or top (text) of the treaties, not immediately next to the text to which they are related as in the original format. Placing the notes at the top of the treaty does give the reader an overview or summary of the treaty at a glance. Frames would have been an ideal way to view the treaties in this project, as the side notes would remain fixed while the text is scrolled and vice versa, but frames were not standard HTML until HTML v3.2 was issued as a recommendation by the World Wide Web Consortium in January 1997. The group was working with HTML v2.0 when the project was begun in the fall of 1996. It was important to work within the HTML standard as advanced browser capability at the time was based on the current standard. The group also considered placing the notes next to the articles to which they referred, as they appear in the original text, using a table within a table technique. But this also was not standard HTML. Again, the design was an attempt in every way to maintain the integrity of the work. Errors in the original text have not been corrected.

Table of Contents

Evaluation

The project was completed in December 1997. Aside from two team members leaving, the project had progressed well and the students had done excellent work. In light of the original proposal, the following paragraphs describe how the objectives were met, future approaches to be taken, and in the next section, the conclusions / recommendations the group reached concerning the continuation of the project.

(1) The first objective was to develop an innovative methodology for reproducing on the WWW text containing complex formats including margin annotations, variable type fonts and tabulated and Columnar information. The design of the treaty layout in table format is unique but uses standard methodologies. After the treaty header, Kappler's notes are in a table row, (width of 20%), followed by the text of the treaty in a second table row (80%). Any text within the treaty that was italicized or in bold was tagged with the appropriate HTML tags. Smaller fonts were preserved by using the FONT SIZE tag. The appearance of all witnesses, signatures and the index were mostly maintained by enclosing these in PRE tags. In the original text, signatures and witness signatures are in alignment. In some of the treaties in table format, this alignment is slightly off due to table structure and the consequent need to reduce the font size. Paragraphs are also not indented.

The proposal was written to encourage the use of HTML to replicate the text of a paper publication. The environment in which the project was accomplished included the following factors: large groups of users were using text only browsers; HTML v2.0 (the standard at the time) was relatively restrictive; advanced browsers were still developing; funding and time were limited. Given these factors and the nature of the proposal, the group's overall approach to this project was innovative.

If the proposal had been written more openly, the project group would have been able to approach the project differently, possibly creating a text only markup and images. See also the Conclusion.

(2) The second objective was to use hypertext and full-text indexing to enhance historical or legal documents. Links for Table of Contents, Search, Index and Pages are at the top of each treaty in both formats. The table of contents and search function are value-added capabilities. The Table of Contents offers access to treaties by tribe. The Search function will allow users to search for keywords in the treaties, table of contents and index. The Index is hyperlinked, allowing users to move from the Index to the treaties very quickly. In addition, Kappler's notes are hyperlinked and

are placed at the top left (table) or top (text) of the treaties. Users can view these treaty highlights for quick access to important text.

(3) The third objective was to determine the cost effectiveness of the developed methodology to digitize the complete set of Kappler's *Indian Affairs: Laws and Treaties* as well as other historical or legal documents. Following is the Average Total Time Per Page for staff and students:

Staff:

HTML markup:

Table: 44.5 minutes per page

Text: 21.5 minutes per page

Avg. Total staff hrs. per page: 66 minutes per page

Students:

Scanning and editing:

Avg. Total student hrs. per page: 32.5 minutes per page

The completion of Vol II (925 pages remaining) would require: 1017.5 staff hours (127 days) and 501 student hours. Paying students \$5.25 an hour and a full-time staff \$30,000 per year, this would cost approximately \$2,630 in student wages and \$30,000 for staff. Objective Computing, a company that has produced Vol II. in CD-ROM format, was contacted and asked about possibly licensing their digitized text if the project would continue. Licensing the text would be a more cost-effective approach.

Completing the seven volumes using the developed methodology would require 3.5 years at a cost of approximately \$130,000. Considering how technology and user needs have changed, if the project were to continue it would be pursued by using the already digitized text of Vol II. and by contracting out for digitization of the remaining volumes. A different methodology would also be employed: tagging the treaties as text-only for all browsers with scanned images of the text included. See the Conclusion.

(4) The fourth objective was to make available on the WWW a faithful representation of an historically significant and long out-of-print resource to the benefit of Native Americans, historians, students and the legal community. This has been accomplished with the digitized treaties in table format. The text version, again, serves users regardless of browser type.

The OSU Library's approach was appropriate for early 1996 and available resources. As will be discussed in the conclusion / recommendations, if the project continues, a different approach will be taken due to the technological advances that have taken place. The team concluded at the end of the project that because of the less than 100% accuracy of OCR and the human eye, images of the pages should additionally be scanned as images and included so that users will have access to searchable text, plus access to an exact replication of the original. As an immediate follow-up to the project, the team will be adding images during January 1998.

(5) The fifth objective was to enhance the skills and knowledge of the project participants relative to presenting information on the WWW. This has been accomplished. Members have benefited greatly by applying technologies at hand and available resources to publish the treaties of the Five Tribes from *Indian Affairs: Laws and Treaties* on the Web. Members have researched currently available technologies and have discussed the best possible way to proceed with the digitization of *Indian Affairs: Laws and Treaties*.

(6) The sixth objective was to attract additional private funding based on the strength of the successfully completed project. The project team will submit a proposal in January 1998 to a

large, private foundation requesting funding to continue the digitization of the seven volumes of *Indian Affairs : Laws and Treaties* .

Contingent on available funding, the Library would like to continue the digitization of *Indian Affairs : Laws and Treaties*. Several approaches have been considered.

Major digitization projects worldwide are now using the Text Encoding Initiative (TEI), an international cooperative research effort, the goal of which is to define a set of generic guidelines for the representation of textual materials in electronic form. By providing a description of information which is independent of realization or media, the TEI scheme (like other Standard Generalized Markup Language-based approaches) enormously facilitates the construction and exploitation of multimedia technology. SGML (ISO 8879) has been chosen as the most appropriate vehicle for the Guidelines. SGML supplies a formal notation for the definition of generalized markup languages. This is the approach that the team recommends the project take if continued.

Adobe Systems, Inc. produces software (Adobe Capture), that creates a digital image which is also searchable. Capture offers a key means of preserving the documents while allowing them to be searched. There are concerns, however, about how Capture handles hand-set type, its longevity (as with any software -will users be able to read it 30 years from now?), and the requirement that certain software (Adobe Acrobat Reader) be installed on a user's computer in order to view file document (although browsers are now coming equipped with Acrobat Reader). This is another possible approach but not as strong due to the outlined weaknesses.

Table of Contents

Conclusion

As the proposal was written, the digitization of 150 pages of complex text from Volume II (Treaties) of *Indian Affairs: Laws and Treaties*, maintaining as much as possible the appearance and intent of the original work, has been successfully completed. The proposal was written to encourage the use of HTML to replicate the text of a paper publication. Other factors in the spring of 1996, when the project was started, include: users using text only browsers, HTML v2.0 was relatively restrictive, advanced browsers were still developing, and funding was limited as was time. The project resulted in the creation of two versions of the treaties: one text-only version that could be read by all browsers, the other in table format, to fulfill the requirements of the proposal.

Using a Hewlett-Packard Scanjet 4C scanner running Caere's Omnipage Pro software v6.0 for scanning the pages using OCR, the text, once the scanner and student were trained, was relatively easy to scan. Initial editing took place using Omnipage editing features. The detailed editing and proofreading of the treaties was time-consuming due to the signatures and inherent spelling errors of the original text. The HTML markup of the table format proved to be difficult and time-consuming because of the signatures in columns, consequent use of the PRE tag within a table format, and linking of notes. Average staff time per page was 66 minutes; average student time per page was 32.5 minutes.

The project is running on the OSU Library's server, a Gateway 2000 P5-200 workstation running \Windows NT v4.0. By early January 1998, the server will be running Microsoft's Internet Information Server (IIS) v4.0 for serving the Web pages and the included Microsoft Index Server v2.0 for indexing the pages. The project URL is: <http://www.library.okstate.edu/kappler/>

The team concluded at the end of the project that due to the less than 100% accuracy of OCR and the human eye, images of the pages should additionally be scanned and included so that users will have access to searchable text plus access to an exact replication of the original. As an immediate

follow-up to the project, the team will be adding images during January 1998. If the proposal had been written more openly, the team would have been able to approach the project differently, creating a text only markup plus images for example. The team also questioned whether the effort to tag text to replicate exactly the original format was worthwhile. Too much time was required. Again, users would benefit as much from text tagged in simple text format plus images.

If the team is able to secure additional, private funding, the OSU Library hopes to complete Vol. II and continue to digitize the remaining six Volumes of *Indian Affairs: Laws and Treaties*. Since technology has advanced considerably in the last two years as has understanding of digitizing text in general, there are greater options for improved access. The project team still agrees that a format of the digitized text should be compatible with all browsers and that digitized text comply with the HTML standard recommended at the time.

The Library has submitted a proposal to an external funding source to continue the digitization of *Indian Affairs: Laws and Treaties*. The project team recommends that the text be tagged in SGML in a simple text format following TEI Guidelines with scanned images of tile text included. A search feature would be included.

The team also recommends to the OSU Library Administration that the project participants attend the spring of 1998 workshop : Digital Imaging for Libraries and Archives (<http://www.library.cornell.edu/preservation/digital.html>), a week-long workshop on the use of digital imaging technology in libraries and archives, sponsored by the Cornell University Library Department of Preservation and Conservation, March 22-27, 1998, in Ithaca, New York.

The OSU Library and the project team wish to thank AMIGOS for their generosity and support for this explorative endeavor .

S. Holcombe, Principal Investigator

December 1997.

Table of Contents

Project Participants

Project Team: Suzanne Holcombe (principal investigator), John Berry, Christine Mueller (now at the University of New Mexico Libraries), Dan Tan (now at Broward County Community College, Ft. Lauderdale, FL).

Project Participants, Students: Michelle Lilburn (scanning and editing), Sarah Segall (proofreading, editing), Stephanie Brown (scanning of images), Janey Williams, Tom Marshall, Deanita Sitton (assistance with Index).

Project Consultants: Sheila Johnson, Jennifer Paustenbaugh, Richard Page. Brian Rossmann, John Phillips, Vicki Phillips.

Project Assistants: Sherry Fletcher, Dan Chaney, OSU Library Systems Department.

Scanner Consultants: Justin Beightol, Jason Razein (Computing and Information Services), Tom LaBron, Sherry Fletcher .

Questions and comments to:
OSU Library: lib-web@okway.okstate.edu